Future Network Requirements for Science

Eli Dart, Network Engineer

ESnet Network Engineering Group

Terabit Networks for Extreme Scale Science Workshop

Washington, DC

February 16, 2011

Overview

Virtual circuits overview

Case studies

- Traffic engineering for the LHC
- Remote data analysis at JGI

Implications for next-generation infrastructure

Virtual Circuits

Virtual circuits provide a virtual link between two points in a network

- Use of the virtual circuit is exclusive to the devices connected to the ends of the circuit, e.g. routers, hosts
- Typically traffic engineering is applied
  - Bandwidth guarantees
  - Specific path through the network

In order to be useful, virtual circuits must in general traverse multiple administrative domains – however, this is a big advantage

- Science is inherently multi-domain – collaborations use resources at multiple institutions, and expect those resources to work together seamlessly
- Very powerful – two hosts on different continents can appear to each other as if they are directly connected, with a guarantee of service

Case Study – Large Hadron Collider (LHC)

Large accelerator experiment

- Two experiments with very large data volume – ATLAS, CMS
- Large scale collaboration – multi-continent, 1000s of physicists
- Data analysis is too large-scale to put in a single location

Large-scale automated data distribution (many PB/year)

- Sheer scale of data, computation, collaboration, etc. requires significant automation
- In order to support large-scale automation, several things are required
  - Network must be well-understood (consistent behavior, reliable)
  - Network must be high-performance (sufficient bandwidth, zero loss)
  - Effective software infrastructure must be widely deployed

Case Study – Large Hadron Collider (LHC)

Virtual circuits on the network play a significant role

- Bandwidth and service guarantees per circuit
- Congestion on best-effort network does not impact VC traffic
- Traffic engineering capabilities provide diversity
  - Multiple paths defined a priori
  - Model proven in production in live multiple-failure scenarios

LHC experiments use large-scale Grid software deployments for data transfer and analysis

- Significant investment in tools
- Large payoff in terms of automation, productivity

Case Study – Remote Data Analysis (JGI)

Large genome sequencing infrastructure

- PB/year generated and analyzed
- Analysis of samples from other institutions
- Future analysis of data from other institutions

Use virtual circuits to attach JGI sequencing facilities to NERSC computation and storage facilities

- Layer2 circuits allow NERSC and JGI resources to function as if they were on the same network
- Remote filesystem mounts
- Recent statistics – 1PB data transferred in one month

LHC and JGI Cases – Analysis

Contributing factors to LHC success

- Internal sophistication (network-savvy, good internal support structures within collaborations)
- Investment in software tools
  - Development
  - Deployment
- These qualities enable effective use of resources

Contributing factors to JGI success

- Experts did the integration (supercomputer center staff)
- Available tools well-suited to the problem (supercomputer center resources, virtual circuit infrastructure)

Coming Challenges

Majority of collaborations do not possess the attributes that made the LHC and JGI cases successful

- Generalized means for driving distributed resources is not available to most collaborations
  - Tools unavailable or poorly suited
  - Advanced network resources (e.g. virtual circuits) do not reach facility, data movers, etc.
- Experts in data movement not present at institutions or in collaboration

Many collaborations will need to implement multi-site infrastructures to support next-generation experiments

- Streaming data between facilities and supercomputer centers
- Near-real-time analysis
- Coupled or co-scheduled runs

Necessary Toolset – Science Services

Extreme scale resources will be complex, few in number

- Effective tools to drive them must be available
- Remote users, data, analysis will exist in the general case, therefore these use cases are part of the core mission
- Tools must run at scientist's site and at exascale site

A generalized set of services must be available in order for the science community to effectively utilize extreme scale resources

- Data movement
  - Correct, consistent, reliable
  - Tools must be available and useful without hero efforts by experts
- Multi-site, multi-domain scheduling and connections
  - Co-scheduling of resources
  - Easy-to-use service interface to integrate Network, VMs, data sources, etc.

Data Integrity

Cautionary note

- Silent corruption of data is already an issue
- TCP checksum is inadequate to guarantee data integrity at large scales (according to one study, Stone and Partridge, one undetected error can be expected every 15TB to 90TB)
- This is simply the wrong order of magnitude for extreme scale data sets

General solution needed

- Several disciplines already take steps to ensure data integrity
- Difficult, time-consuming – “it’s a pain” – but it must be done
- Next-generation tools and services must consider “always-on” integrity mechanisms

Questions?

Thanks!